BOIES SCHILLER FLEXNER LLP
David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsfllp.com

**LIEFF CABRASER HEIMANN &
BERNSTEIN, LLP**
Rachel Geman (*pro hac vice*)
250 Hudson Street, 8th Floor
New York, New York 10013-1413
(212) 355-9500
rgeman@lchb.com

**CAFFERTY CLOBES MERIWETHER &
SPRENGEL LLP**
Bryan L. Clobes (*pro hac vice*)
135 S. LaSalle Street, Suite 3210
Chicago, IL 60603
(312) 782-4880
bclobes@caffertyclobes.com

*Counsel for Individual and Representative
 Plaintiffs and the Proposed Class*

(additional counsel included below)

JOSEPH SAVERI LAW FIRM, LLP
Joseph R. Saveri (SBN 130064)
601 California Street, Suite 1505
San Francisco, California 94108
(415) 500-6800
jsaveri@saverilawfirm.com

DICELLO LEVITT LLP
Amy Keller (*pro hac vice*)
10 North Dearborn Street, Sixth Floor
Chicago, Illinois 60602
(312) 214-7900
akeller@dicellolevitt.com

Matthew Butterick (SBN 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
(323) 968-2632
mb@butticklaw.com

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

RICHARD KADREY, et al., *Individual and Representative Plaintiffs*, v. META PLATFORMS, INC., *Defendant.*	Case No. 3:23-cv-03417-VC **PLAINTIFFS' NOTICE OF MOTION AND MOTION FOR PARTIAL SUMMARY JUDGMENT**

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page(s)

Cases

Other Authorities

NOTICE OF MOTION AND MOTION

TO THE COURT, ALL PARTIES, AND THEIR ATTORNEYS OF RECORD:

PLEASE TAKE NOTICE that on May 1, 2025, at 10:00 a.m., or as soon thereafter as the parties may be heard, before the Hon. Vince Chhabria, District Judge, U. S. District Court for the Northern District of California, in Courtroom 4 – 17th Floor, 450 Golden Gate Ave., San Francisco, CA 94102, Richard Kadrey, Sarah Silverman, Junot Diaz, Andrew Sean Greer, David Henry Hwang, Matthew Klam, Laura Lippman, Rachel Louise Snyder, Lysa TerKeurst, Jacqueline Woodson, and Christopher Farnsworth (collectively, "Plaintiffs")[1] will and hereby do move this Court for an order granting partial summary judgment against Defendant Meta Platforms, Inc. ("Meta").

Plaintiffs seek an order pursuant to Federal Rule of Civil Procedure 56 granting their Motion for Partial Summary Judgment, on the grounds that summary judgment is warranted because uncontroverted evidence establishes: (1) Meta committed direct copyright infringement under 17 U.S.C. § 501 of each of Plaintiffs' Copyrighted Books asserted in the Third Amended Consolidated Complaint, Dkt. 407; and (2) Meta's reproduction of Plaintiffs' Copyrighted Books without permission, including through peer-to-peer file sharing, is not fair use under 17 U.S.C. § 107.

Plaintiffs' Motion for Partial Summary Judgment ("Motion") is based on this Notice of Motion and the accompanying Motion, all pleadings and papers in this action, and oral argument of counsel.

---

[1] Plaintiffs Ta-Nehisi Coates and Christopher Golden do not seek summary judgment at this time.

MEMORANDUM OF POINTS AND AUTHORITIES

Under a straightforward application of existing copyright law, Meta is liable for massive copyright infringement. By taking Plaintiffs' books—along with millions of other copyrighted works—from pirated online databases,[2] Meta "g[o]t for free something [it] would ordinarily have to buy." *A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1015 (9th Cir. 2001), *as amended* (Apr. 3, 2001) (affirming order that "downloading MP3 files does not transform the copyrighted work" and is not fair use). And to more easily obtain this copyrighted data, Meta torrented it, using a process by which Meta made available and distributed the copyrighted data to other online pirates.

Meta could and should have paid to buy and license literary works from copyright holders to train its Large Language Models ("LLMs"), named "Llama." To train Llama to mimic human expression when producing text output, Meta decided it needed a large corpus of high-quality text, particularly books. Meta wanted books for their expressive content—the very subject matter copyright law protects. But instead of paying rightsholders, Meta systematically took and fed entire copies of pirated works into its LLMs to extract that expressive content without having to pay.

Whatever the merits of generative artificial intelligence, or GenAI, stealing copyrighted works off the Internet for one's own benefit has always been unlawful. *UMG Recordings, Inc. v. MP3.Com, Inc.*, 2000 WL 710056, at *1 (S.D.N.Y. June 1, 2000) (the "mere fact" that copyright infringement is "clothed in the exotic webbing of" a new technology "does not disguise its illegality") (Rakoff, J.). Meta knowingly used pirated databases to copy massive quantities of copyrighted works—all of Plaintiffs' Copyrighted Books, books written by hundreds of thousands other authors, and even books authored by at least 10 Supreme Court justices who served in this century, including Justice Breyer's *Making our Democracy Work* and Justice Ginsburg's *My Own Words*.[3] These pirated databases are illegal, routinely targeted by government enforcement

---

[2] Meta's torrenting expert produced workpapers cataloguing the works Meta torrented from pirated databases in 2024 alone, which reveal 9.7 million files from Z-Library, 5.8 million files from Internet Archive, and 489,000 files from LibGen Non-Fiction. *See* Exs. 93-95. The text files are far too large for e-filing, so Plaintiffs instead will provide them to the Court via digital courtesy copy.

[3] Sourced from Ex. 93, meta_ia_downloads.

PLAINTIFFS' NOTICE OF MOTION AND MOTION FOR PARTIAL SUMMARY JUDGMENT
CASE NO. 3:23-cv-03417-VC

agencies for criminal and civil copyright infringement. Authorities regularly shut down their domains and even prosecute the perpetrators. That Meta knew taking copyrighted works from pirated databases could expose the company to enormous risk is beyond dispute: it triggered an escalation to Mark Zuckerberg and other Meta executives for approval. Their gamble should not pay off.

There are aspects of this case that belong to a jury, including Meta's scienter in stripping copyright management information from millions of copyrighted works to conceal infringement in violation of the DMCA; whether fair use applies to Meta's infringements *during and after* the LLM training process; and issues that impact ultimate remedies. But due to the undisputed material facts, Plaintiffs are entitled to summary judgment now with respect to two issues.

First, Plaintiffs are entitled to summary judgment on direct copyright infringement. To establish infringement, a plaintiff must demonstrate that the defendant violated one of the exclusive rights protected under the Copyright Act by showing (1) ownership of a valid copyright; and (2) copying of constituent elements of the work that are original. Here, there is no genuine dispute of material fact for the moving Plaintiffs. Meta infringed each of their copyrights, full stop.

Second, Plaintiffs are entitled to summary judgment on the ground that Meta's initial acquisition of millions of pirated works cannot be fair use. There is no dispute that Meta torrented a minimum of ***267.4 terabytes*** ("TB") of works from known pirated databases to further its GenAI efforts. There is no dispute that Meta torrented at least *666 copies* of Plaintiffs' copyrighted books in addition to millions of other copyrighted works. And there is no genuine dispute that Meta made widely available and even *reuploaded* to other online pirates at least some quantity of that pirated data as part of the peer-to-peer ("P2P") sharing process. Meta's response in this case seems to be that a powerful technology corporation should not be held to the same standard as everyone else for illegal conduct. That its theft was so massive, it must somehow be "fair use." Call it the Bob Dylan defense: "Steal a little and they throw you in jail / Steal a lot and they make you king." Bob Dylan, "Sweetheart Like You" (Columbia Records 1983). But there is no-fair use defense to Meta's unmitigated piracy of copyrighted works, let alone Meta's acquisition of them through

torrenting, a P2P file sharing process that "obvious[ly]" "does not constitute fair use." *In re DMCA § 512(h) Subpoena to Twitter, Inc.*, 608 F. Supp. 3d 868, 879 (N.D. Cal. 2022) (Chhabria, J.).

FACTUAL BACKGROUND

I.   THE PARTIES

A. Author Plaintiffs

Plaintiffs[4] are book authors with registered copyrights ("Copyrighted Books" or "Books"). Ex. 1, Plaintiffs' Copyright Registrations.[5] As owners of these registered copyrights, Plaintiffs own the exclusive rights to reproduce their own Copyrighted Books. 17 U.S.C. § 106.

There are no material disputes about the following facts. ***First**,* Meta took Plaintiffs' Copyrighted Books—which were included in pirated datasets—without paying for them.[6] ***Second**,* Meta used Plaintiffs' Copyrighted Books as AI LLM training data, without approaching Plaintiffs, their agents, or their licensing intermediaries to license their Copyrighted Books for these purposes. Ex. 2, Meta RFA Compilation, No. 9. ***Third**,* Meta did so notwithstanding the fact that Plaintiffs are amenable to such licensing. Ex. 4, Plaintiffs' RFA Compilation.

B. Defendant Meta Platforms, Inc.

Meta is a $1.67 trillion conglomerate that owns and operates—among other products—Facebook, Instagram, Threads, and WhatsApp. Meta develops, markets, and sells its LLMs, the Llama models, as standalone commercial products and as key components integrated into its other commercial products. Ex. 5, Rao Tr. 69:16–70:15, 136:10–136:13, 178:22–179:6 (explaining Llama 3 has a commercial benefit and Llama has been embedded into Meta's other commercial products). Meta generates revenue from its Llama models and will continue earning revenue from each iteration. Ex. 6, Meta ROG Compilation, No. 20 (explaining Meta has earned money from

---

[4] Plaintiffs Christopher Golden and Ta-Nehisi Coates do not seek summary judgment at this time.

[5] All Exhibit citations are to the accompanying declaration of Maxwell V. Pritt.

[6] Ex. 2, RFA Nos. 9 (admitting Meta did not compensate Plaintiffs for Copyrighted Books); 45–89 (admitting substantially all the text of Plaintiffs' (except for Plaintiff Farnsworth's) Copyrighted Books are in Books3 datasets that Meta acquired); 90–91, 94, and 96 (admitting Meta used a dataset including substantially all of Plaintiff Farnsworth's Copyrighted Books); Ex. 3, Summary Table of Copies (showing all Copyrighted Books were in the pirated datasets Meta downloaded and later incorporated into its LLM training datasets).

agreements with ████████████████████ in which "Meta shares a percentage of the revenue that they generate from users of the Llama Models . . . hosted by those companies"); Ex. 7, Meta_Kadrey_00093345 (revenue spreadsheet for Llama 2 and 3). In June 2024, Meta estimated that its 2025 GenAI-driven revenue will be more than $2 billion. Ex. 8, Meta_Kadrey_00089020, at -020 (estimating $2-3 billion of GenAI revenue in 2025 with $460 billion ("Base Case") to $1.4 trillion ("GenAI Wins Case") of Total Revenue through 2035).

Meta hopes to use its Llama models to become a leader in GenAI.[7] Meta has made, and plans to continue to make, significant investments in developing its AI. Ex. 12, Meta_Kadrey_00157008 (2024 GenAI budget showing over $900 million of Operating Expenses with plans for 2025 Operating Expenses exceeding $1 billion). In fact, Meta's total 2025 AI investments, including infrastructure, are expected to exceed $60 billion.[8]

II.   META HIGHLY VALUES BOOKS FOR LLM TRAINING AND DEVELOPMENT

Meta's initial foray into GenAI was rocky. Early iterations of Meta's Llama models were viewed as primitive research projects that lagged far behind industry leaders. Meta knew Llama did not perform to the level of its competitors. Ex. 13, Meta_Kadrey_00150660 ("lagging behind ChatGPT"); Ex. 14, Meta_Kadrey_00045402, at -418 ("We are significantly lagging [OpenAI and Google]"). To catch up, Meta focused on procuring high-quality long form text data.[9]

A. Meta Quickly Identified Books as High-Quality LLM Training Data.

Early in the process of developing the Llama models, Meta's AI researchers identified literary works, especially books, as a key source of high-quality, long-form text data. In October

---

[7] Ex. 9, Meta_Kadrey_00046383 ("Our research vision presents a unique opportunity to be the leader of Generative AI by taking a different . . . approach than the rest of the industry, namely by being open and by building a strong research and developer community around our technology."); Ex. 10, Meta_Kadrey_00044759, at -760 ("We see Llama as a competitive advantage and value flowing back to Meta"); Ex. 11, Meta_Kadrey_00046203 ("With the increased research advancement in LLMs, we want to continue to position FAIR as a leader in this space.").

[8] Bruna Horvath, *$60 billion in one year: Mark Zuckerberg touts Meta's AI investments*, NBC News (Jan. 24, 2025), https://www.nbcnews.com/tech/tech-news/60-billion-one-year-zuckerberg-touts-metas-ai-investments-rcna189131.

[9] Ex. 15 ("Goals: . . . Get as much long form writing as possible in the next 4-6 weeks"); Ex 16, Meta_Kadrey_00171954 ("Hence, continuing pretraining will be effective only if we have relevant self-supervised data, like books").

2022, a Meta research engineer informed senior Meta engineers that "[i]f data quality and volumes are so important for the performance of LLMs, the best resource we can think of are definitely books." Ex. 18, Meta_Kadrey_00074729. Meta believed books were high quality training data for LLMs because they generally are well-written representations of human language, with a coherent structure and strong grammar. Ex. 19, Meta_00154729 ("Books are generally high quality text."). Meta understood that the higher quality of the text, the better it could train its Llama models. As Meta's director of product for GenAI, Eugene Nho, testified, "not every words [sic.] are equally insightful, or helpful, to teach the model." Ex. 20, Nho. Tr. 49:19–49:20.

To compete with other LLMs like ChatGPT, Meta also understood that it needed book data to develop "long context windows"—that is, the ability to produce output based on long prompts. Ex. 21, LeCun Tr. 49:17–50:7 (explaining why developers need long text data to create long context windows); Ex. 22, Meta_Kadrey_00036819, at -823 ("It's interesting to see how the length of documents in the datamix impacts the long context for training.").[10] Meta knew that books contained expressive content, which could ensure the competitiveness of its LLMs.[11]

B. Technical Experts Agree that Books Constitute High-Quality Training Data.

The parties' experts agree that high-quality text data is imperative for training because it builds diverse and nuanced relationships between words, improving output quality. Ex. 27, Lopes Opening Report, ¶ 75; Ex. 28, Ungar Opening Report, ¶ 51. Dr. Emily Bender, a computational linguist, explains that the value of books lies in "the selection and arrangement of the words contained therein." Ex. 29, Bender Rebuttal Report, ¶ 86. Meta's LLM expert similarly agrees that high quality data definitionally "captur[es] useful semantic relationships." Ex. 28, ¶ 216. Both

---

[10] Ex. 23, Meta_Kadrey_00057170, at 170-71 ("It is becoming increasingly important to research the best strategies for creating long context LLMs . . . Data sources that are naturally longer[:] Books corpus (B3G or newer variants."); Ex. 24, Meta_Kadrey_00173041 ("Anna's Archive is full of long context books that they will desperately need for llama4 to have long context and there was a substantial bump in quality from libgen for similar reasons.").

[11] Ex. 13 ("lagging behind ChatGPT" in creativity and style); Ex. 25, Meta_Kadrey_00150659 ("[Libgen] would probably contribute the most towards improving our Agents capabilities and their creativity."); *see also* Ex. 26, Meta_Kadrey_00093951, at -953 (Meta's Al-Dahle stating "For LLM, we need to focus on nailing creativity and entertainment prompts.").

sides' experts further agree that books are uniquely valuable as data for developing longer-context windows. Ex. 28, ¶ 239 (noting The Pile, which includes Books3, can "effectively extend[] context length indefinitely"); *see also* Ex. 27, ¶ 183 ("To train an LLM for long-context tasks, the training data must be long but consistent—consistent topic, recurrent words, similar style, etc. Books are a good fit for this purpose.").

III.   AS META SCRAMBLED TO CATCH UP WITH ITS COMPETITORS, IT RESORTED TO COPYING PIRATED BOOKS EN MASSE WITHOUT PERMISSION.

A.   Meta Initially Explored Using Pirated Databases But Deemed Many of Them Too Risky.

In 2022, Meta knew its competitors were obtaining superior results largely due to their use of pirated datasets to train their LLM models.[12] So Meta devised a simple strategy to catch up with its competitors: acquire more books—fast.

In October 2022, Meta's AI team began seeking legal approval for "pure exploration work" regarding the performance benefits that could be achieved by training Llama on books and articles obtained from Library Genesis ("LibGen"), a notorious online shadow library of pirated copyrighted works. Ex. 31, Meta_Kadrey_00158651. Since 2015, federal courts have consistently found that LibGen violates U.S. copyright law. *Elsevier Inc. v. www.Sci-Hub.org,* 2015 WL 6657363 (S.D.N.Y. 2015); *see also Cengage Learning, Inc. v. Does 1-50 d/b/a Library Genesis*, Case No. 23-cv-8136-CM, Dkt. 36 at 2-3 (S.D.N.Y. Sept. 24, 2024) (granting permanent injunction against LibGen for copyright infringement).

At first, Meta's LibGen plan was to "use it just for pure exploration to see if there is value" for model performance, and if so, "then we will setup proper licensing agreement." Ex. 32, Meta_Kadrey_00218170. Other researchers affirmed that "we need to be very careful to use only licensed data for the actual training of the model." *Id.* In late October 2022, Meta's in-house counsel gave the "green light to use [LibGen] for pure exploration." Ex. 33,

---

[12] Ex. 18 ("everyone is using lib-gen (startups, but also google, openAI)"); *id.* ("And I'm pretty sure other folks have no issues taking all of libgen 😊"); Ex. 30, Meta_Kadrey_00172691 (predicting that Anthropic found "4x more books" than Meta through "libgen prolly").

Meta_Kadrey_00159323. Once exploratory approval was granted, Guillaume Lample, a then-Meta engineer and now co-founder of AI competitor Mistral AI, likely torrented Meta's first copy of LibGen. Ex. 34, Clark 30(b)(6) Torrenting Tr. 54:14-25. That copy of LibGen has since vanished from Meta's possession, but employees within Meta speculated it was acquired via torrenting, as discussed further *infra. Id.* at 55:23-56-5.

When finalizing Llama 1, Meta decided that the "exploratory" LibGen data was too risky to use in an open-source model.[13] But Meta still used pirated books from "The Pile," a popular dataset developed by EleutherAI for the very purpose of training LLMs.[14] Specifically, Meta relied on Books3, a subset of The Pile that contains just under 200,000 long-form works, including all of Plaintiffs' Copyrighted Books. Ex. 2. Meta disclosed its use of Books3 in the Llama 1 Paper,[15] but for later Llama models, Meta made a concerted decision to conceal the fact that it was continuing to train on Books3 because it comprised pirated books. While the controversial contents of Books3 only became publicly known through an August 2023 exposé published in *The Atlantic*, Meta had long known Books3 had "issues with copyrighted works." Ex. 37, Meta_Kadrey_00054898. When Meta received an advance copy of the *Atlantic* exposé one day before its publication, employees worried that the public would realize Meta was continuing to use pirated data for training: "[i]t's the piracy (and us knowing and being accomplices) that's the issue." *Id.*; *see also* Ex. 38, Meta_Kadrey_00054433 ("ah yikes books3 contains more than I think the lawyers realize"); Ex. 39, Esiobu Tr. 31:5-32:10.

B. Meta Briefly Tried To License Books But Quickly Abandoned Those Efforts.

After the release of Llama 1 in February 2023, Meta's engineers pushed to gather high-quality long-form text data for training subsequent iterations. Ex. 15, Meta_Kadrey_00232214

---

[13] Ex. 35, Meta_Kadrey_00066092 ("The team was originally planning on an internal-only model due to libgen constraints. The new data mix would allow an open source release, and the team would like to plan for it."); *see generally* Ex. 36, Meta_Kadrey_00063689 (discussing experiments for what would become Llama 1 using "proprietary" datasets such as LibGen).

[14] Leo Gao et al., *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*, ELEUTHERAI (Dec. 31, 2020), available at https://arxiv.org/abs/2101.00027.

[15] Hugo Touvron et al., *LLaMA: Open and Efficient Foundation Language Models*, META AI, at 2 (Feb. 27, 2023), available at https://arxiv.org/abs/2302.13971.

("Goals: (from Ahmad [Al-Dahle]) . . . Get as much long form writing as possible in the next 4-6 weeks . . . Books—all genres."). Melanie Kambadur, a senior manager and researcher, stressed that same month that it was "really important for [Meta] to get books ASAP" because "books are actually more important than web data." Ex. 40, Meta_Kadrey_00233188.

To close the gap between Llama and its competitors, Meta pursued licensing deals with publishers in March and April 2023.[16] These same publishers published many of Plaintiffs' Copyrighted Books.[17] Importantly, Meta knew that the publishers' books had significant value as AI training data, and for a time, Meta was prepared to pay millions to license them. In January 2023, Meta's Business Development ("BD") team was given a **$17 million** budget for "dataset licensing," and by March, Meta requested to increase that budget to **$27 million**. Ex. 48, Meta_Kadrey_00207194, at -195. As deals with major publishers moved closer to fruition, the budget increased: "$50M would prolly be reasonable." Ex. 49, Meta_Kadrey_00233332. By April, Meta discussed a **$200 million budget** for training data, including "**100 of our 200M for books.**" Ex. 92, Meta_Kadrey_00235020, at -029-030. As Ahmad Al-Dahle, VP and Head of GenAI put it, "It would be a travesty if our models suck because we didn't spend $100M on data." *Id.* at -030.

C.   Meta Abandoned Its Licensing Efforts After It Discovered It Could Just Take the Copyrighted Books It Wanted from Pirated Databases Without Paying for Them.

Meta abruptly ended its licensing efforts on April 7, 2023. Ex. 50, Boesenberg Tr 130:17-132:10; 383:5-384:12. That directive to stop licensing was even "escalated to MZ [Mark Zuckerberg] re: fair use." Ex. 51, Meta_Kadrey_00235350, at -351. From there, Meta's Business

---

[16] Ex. 41, Meta_Kadrey_00234554 ("We should try to get all the big guys: Harper Collins, Simon & Schuster, Macmillan, Hachette, Penguin Random House."). Meta entered into preliminary negotiations with nearly every major book publisher about using their copyrighted works as training data. Ex. 42, Meta_Kadrey_00140688 (emails to Penguin Random House about licensing opportunity); Ex. 43, Meta_Kadrey_00152934 (emails to Harper Collins about licensing opportunity); Ex. 44, Meta_Kadrey_00051952, at -954-55 (strategy document discussing potential licensing deals for fiction with Harper Collins, Simon & Schuster, Macmillan, and Bartelsmann).

[17] Meta even *successfully executed* four licenses in 2022 with certain African-language publishers to use a limited number of individual books for training Meta's LLMs. Ex. 45, Meta_Kadrey_00146968 (licensing three books for $21,000); Ex. 46, Meta_Kadrey_00140774 (one-year licensing agreement with ▮▮▮▮▮▮▮▮▮); Ex. 47, Meta_Kadrey_00142650 (discussing three instances of having licensed books for training data).

Development ("BD") team was verbally instructed to stop all text data licensing efforts immediately. Meta's book licensing efforts have never since resumed in earnest. One Meta engineer summed it up perfectly: *"if we license once [sic.] single book, we won't be able to lean into the fair use strategy."* Ex. 52, Meta_Kadrey_00209134, at -135 (emphasis added).

Meta instead resorted to using pirated versions of copyrighted books acquired from illegal online databases. And to conceal its actions, Meta's employees started referring to these pirated datasets as "external" or "publicly available" instead of "pirated," Ex. 53, Meta_Kadrey_00235485, at -486; Ex. 54, Meta_Kadrey_00101583; although even Meta employees knew it was not accurate to refer to pirated datasets as "publicly available." *See* Ex. 55, Clark (30)(b)(1) Tr. 144:21-148:3. While Meta has now trained its LLMs on data acquired from several pirated websites,[18] the pivotal moment occurred when Meta approved using LibGen in April 2023. LibGen is perhaps the best-known illegal site for accessing pirated books, and it has repeatedly been shut down by federal courts for massive copyright violations. *E.g., Cengage Learning,* No. 23-cv-08136-CM, Dkt. 36 (S.D.N.Y. Sept. 24, 2024); Ex. 39, Esiobu Tr. 101:12-102:1. But LibGen's reputation for notoriety did not deter Meta. In May 2023, Meta's data engineers began acquiring millions of copyrighted works of fiction and other genres from LibGen, as well as scientific textbooks, and approximately 81 million scientific articles. Ex. 56, Meta_Kadrey_00089791. These data engineers were directed to download LibGen in order to assess "to what extent libgen data already has papers/textbooks we could purchase from ███████ ███████" Ex. 57, Meta_Kadrey_00160779. The purpose of this experiment was obvious: Meta wanted to know if it would "still need to buy ███████" or if it could generate comparable results from LibGen without having to pay for the literary works in the ███████ catalogues. Ex. 58, Meta_Kadrey_00235448, at -449. The results showed that "up to 90% of books are present in LibGen for ███████ and up to 76% for ███████" Ex. 56 at -861.00013. Meta thus acquired a free substitute for copyrighted literary works and text data that would otherwise cost millions to license.

---

[18] To date, Meta has acquired troves of copyrighted works without compensation from at least the following databases: Books3, LibGen, Z-Library ("Z-Lib"), and Internet Archive.

See Ex. 59, Sinkinson Tr. 163:12-164:3 (noting in the context of LibGen vs. ███ that "your willingness to pay for any training materials depends on what other training materials you already have."); Ex. 92 at -030 ("Books strategy: libgen [in progress] – FREE").

From there, the necessary approvals were obtained by Ahmad Al-Dahle, who "cleared the path" for using LibGen. Ex. 60, Meta_Kadrey_00048149, at -152. Other Meta documents confirmed that "after a prior escalation to MZ [Mark Zuckerberg], GenAI has been approved to use LibGen for Llama 3 (with VP sponsor requiring to accept full risk)" despite that LibGen was "a dataset we know to be pirated." Ex. 61, Meta_Kadrey_00232241, at -244, -246. The use of pirated works was not widely publicized: senior leadership and engineers knew, but it was only conveyed to others "on a 'need to know' basis." Ex. 62, Meta_Kadrey_00211848. Indeed, several employees expressed strong reservations. One Meta researcher commented, "I feel that using pirated material should be beyond our ethical threshold." Ex. 32 at -171. Another referred to LibGen as an "illegal pirated website[ ]" and expressed that "it should not go in the training of the published model." Ex. 18 at -700. One engineer, Todor Mihaylov, outright asked, "Is libgen this" while sharing a screenshot of a Google search result that read, "No, Libgen is not legal, and using Libgen may open you up to legal ramifications." Ex. 63, Meta_Kadrey_00233325-326. Nevertheless, Meta viewed LibGen as "essential" to meeting its performance targets. Without it, Meta's researchers doubted they could "close [the] gap" with competitors OpenAI and Mistral. Ex. 61 at -244. Meta therefore accepted the legal risks and used LibGen to train Llama 3, protecting itself with a flimsy "mitigation" where Meta would remove any books whose file names "self-identified as being 'stolen' or 'pirated,'" such "pirated copy of some silly title - torrent leak.pdf." Ex. 90, Meta_Kadrey_00235739, at -740-41. Another engineer's proposed "mitigation" of searching the books for copyright headers as evidence of piracy was rejected because "that is normal text that is part of many books preface" within LibGen. *Id.* at -741.

As 2023 progressed and Meta's models increased in size, Meta's hunger for high-quality data increased. In late 2023, as development started for Llama 4, Meta's engineers explored the use of Anna's Archive—an aggregator of many pirated datasets. And in March 2024, Meta began

a project to "download/process anna's archive": "essentially a bigger libgen[]." Ex. 64, Meta_Kadrey_00158387. Despite Meta engineers describing Anna's Archive as "a pretty shady website :-P" that "won't be popular with the lawyers," Ex. 65, Meta_Kadrey_00232764, it was used anyway because it aggregates several pirated databases and showed promise for improving Llama's performance. *Id.* Meta engineers confirmed that Anna's Archive contained substantially all of LibGen, nearly all of pirated database Z-Library ("Z-Lib"), and over two-thirds of pirated database Internet Archive. Ex. 66, Meta_Kadrey_00161213, at -214-15. These new datasets dwarfed LibGen's size. An intermediate update posted by Xiaolan Wang showed that by mid-April 2024, Meta had already downloaded 46 TB of pirated data from Internet Archive; 25.7 TB from Z-Lib; and 10 TB from LibGen. Ex. 17, Meta_Kadrey_00107954.

D. To Expedite the Pace of Acquisition, Meta Resorted To Torrenting Massive Quantities of Copyrighted Works.

As Meta began acquiring terabytes of pirated copyrighted works, it quickly ran into a technical issue: downloading such large quantities of files posed an immense strain on Meta's networks and proceeded very slowly. According to Meta engineer Nikolay Bashlykov's notes:

> "Trying to load [LibGen] scimag with the same approach (direct download) as before – doesn't seem to be fast (250k docs / 12h -> 160 days to download the library). Exploring other options to load faster."

Ex. 56 at -835.

Enter torrenting. BitTorrent is a "protocol . . . that is used to distribute a large computer file (such as of digitized music or video) that has been segmented in small pieces between a large number of peer-to-peer [P2P] users."[19] In most cases, and in this case too, users who download via torrent also upload the same files they are downloading to reap the benefits of faster file sharing. In this regard, torrenting is a P2P system where "the *downloaders* of a file *barter* for chunks of it by uploading and downloading them in a tit-for-tat-like manner to prevent parasitic behavior."[20] *See also* Ex. 67, Barbara Frederiksen-Cross ("BFC") Tr. 130:7-131:4 (discussing same).

---

[19] *Torrent*, MERRIAM-WEBSTER.COM, https://www.merriam-webster.com/dictionary/torrent.

[20] Johan Pouwelse et al., *The Bittorrent P2P File-Sharing System: Measurements and Analysis*, IPTPS 2005, at 206 (2005) (emphasis in original).

Moreover, this "tit-for-tat" exchange "optimize[s] the effectiveness and the speed of transfer" of data, as peers prioritize data sharing with peers that are sharing back. Ex. 67, BFC Tr. 130:16-24. In other words, and as Meta's own expert testified, the most effective way to acquire data through P2P sharing is if you are uploading while simultaneously downloading. *Id.* Fundamentally, there exist two phases of data uploading in the torrent process. The first, called "leeching," involves the simultaneous "tit-for-tat" reuploading that occurs during downloading. The second, called "seeding," occurs after a user completes downloading a data file but continues to offer or "seed" the data file to other users.[21] As multiple studies have found, as much as 99% of all data shared via BitTorrent consists of copyright infringing material.[22]

Meta's engineers were well aware of the legal risks posed by pirating data through P2P sharing, but they decided to do it anyway, apparently with the approval of Meta's in-house counsel. Ex. 34, Clark 30(b)(6) Torrenting Tr. 76:5-7; 99:23-100:3. In April 2023, Todor Mihaylov remarked that "if we are certain we want to go this way, we can find pretty much every book that exists – in torrents 😉." Ex. 63. But Meta engineers initially viewed torrenting as off-limits due to its obvious illegality. Ex. 56 at -856 ("using torrents would entail 'seeding' the files – i.e. sharing the content outside, this could be legally not OK"). Indeed, Mihaylov even told his colleagues, "Btw, it would not be trivial to download libgen if everything is in torrents," sharing a link to a Quora webpage asking, ***"What is the probability of getting arrested for using torrents in the USA?"*** Ex. 58 at -451. Nikolay Bashlykov separately wrote, "not sure we can use meta's IPs to load through torrents pirate content" before remarking, "I think torrenting from a corporate laptop doesn't feel right 😅." Ex. 68, Meta_Kadrey_204223, at -223-2. As Meta's expert testified, one

---

[21] *See* Izal et al., *Dissecting BitTorrent: Five Months in a Torrent's Lifetime*, INSTITUT EURECOM, § 2 (2004) ("BitTorrent distinguishes between two kinds of peers depending on their download status: clients that have already a complete copy of the file and continue to serve other peers are called seeds; clients that are still downloading the file are called leechers.").

[22] Jacqui Cheng, *BitTorrent census: about 99% of files copyright infringing*, ARSTECHNICA (Jan. 29, 2010), https://arstechnica.com/information-technology/2010/01/bittorrent-census-about-99-of-files-copyright-infringing/; Jacqui Cheng, *Only 0.3% of files on BitTorrent confirmed to be legal*, ARSTECHNICA (July 23, 2010), https://arstechnica.com/tech-policy/2010/07/only-03-of-files-on-bit-torrent-confirmed-to-be-legal/.

risk with using a "Meta IP" would be that Meta's torrenting could be detected. Ex. 67, BFC Tr. 233:20-25.

In May 2023, Meta's engineers discussed their frustration with the pace of directly downloading LibGen, which was "quite slow" and was "regularly throttled by the server." Ex. 69, Meta_Kadrey_00231275. Bashlykov considered torrenting LibGen, which "would be much faster," but had the same "con": Meta needed to "'seed' the files from our fair cluster servers" to realize any benefits. Ex. 69. Nevertheless, Bashlykov elected to torrent the entirety of LibGen's SciMag section, which held approximately 81 million scientific articles. Ex. 70, Meta_Kadrey_00204235 (admitting he "had to fallback to torrenting since the volume was too large" after engineer Jelmer Van de Linde cautioned, "I can't say I'm particularly comfortable with it from a legal or ethical point of view right now."); Ex. 56 (SciMag contained 80.6 TB). Bashlykov wrote a script that worked to prevent "seeding" the pirated data *after* downloading was complete, but it did nothing to prevent Meta's uploading of the pirated data *while* it was torrenting. *See* Ex. 71, Choffnes Report, ¶¶ 16, 19.

Meta's embrace of data piracy increased exponentially over time. In April 2024, Meta torrented from the shadow libraries available through Anna's Archive, which included data that had been uploaded to LibGen, plus millions of works from Internet Archive and Z-Lib. In so doing, ***Meta became a host and distributor of over 267 TB of pirated material***—a size comparable to the entire printed collection of the Library of Congress twenty times over,[23] or the equivalent of a stack of 2 billion pages of text.[24] Given this enormous quantity of data, and the fact that it took *over two months* to torrent, sometimes on six virtual computers running in parallel, Meta's concerns about the legal risks associated with torrenting were on point. Ex. 34, Clark 30(b)(6) Torrenting Tr. 52:2-25; 100:9-10; *see also id.* 98:16-99:3 (discussing further escalations of legal risks at this stage). Further escalating the risk is the fact that Meta was not merely a passive bystander. Based on the symbiotic role that peers play within a torrent swarm, Meta's torrenting

---

[23] *What is a Terabyte?*, TERADATA, https://www.teradata.com/glossary/what-is-a-terabyte.

[24] Answers given by Meta.AI. *See* Appendix B.

expert, Barbara Frederiksen-Cross ("BFC"), admitted it was "more likely than not" that Meta was "contributing to the overall bandwidth, content, storage, and processing power of that [P2P] network." Ex. 67, BFC Tr. 101:7-104:5. Meta therefore became part of the very "plumbing" that kept the pirated databases functioning well for other users. *Id.* 103:1-2.

Notably, Meta could have changed the default settings on its torrent client to completely eradicate the risk of uploading pirated data while torrenting (i.e., during the "leeching" phase) and "seeding" that data with other peers after the torrented files had been fully downloaded. But it did not do so. Frederiksen-Cross admitted as much.[25] Instead, Meta's only effort to minimize data sharing with fellow pirates was through a script it wrote that aimed to minimize—but not prevent— the amount of pirated data is "seeded" to its peers. Ex. 72, BFC Report, ¶ 100; Ex. 71, ¶ 19. But *while* Meta was torrenting 267.4 TB of pirated data over the course of multiple months and through multiple virtual computers, the unchanged default settings of its torrent client ensured that it was sharing this data with other peers the entire time it was downloading this enormous amount of pirated data. Ex. 71, ¶ 19. Meta's own torrenting expert testified, "I don't know if anyone at Meta possessed that knowledge of libtorrent. I don't even know if they knew there were such parameters." Ex. 67, BFC Tr. 270:2-5. It is hard to believe that Meta software engineers could not have figured out how to effectuate simple changes to their libtorrent client (the answer is one search away on the Internet). The more plausible explanation is that Meta was unwilling to do anything that would slow its data acquisition. And since P2P sharing is optimized for "preferred trading partner[s]," i.e., peers who download and upload simultaneously, changes to these settings were not an option. *See* Ex. 67, BFC Tr. 130:23-24.

Meta opted to hedge this risk by using Amazon Web Services ("AWS") for its torrenting activities specifically, a deviation from Meta's usual practice. When an engineer asked why "FB infra[structure]" could not be used, he was told it was to "avoid[] risk of tracing back the

---

[25] Ex. 67, BFC Tr. 41:1-22 & 44:8-45:15 (noting Meta engineers used default torrent configuration); 107:19-108:18 (admitting the default torrent configuration allows for simultaneous uploading and downloading); 246:13-247:12 & 248:11-23 (noting there are ways to configure the torrent to eliminate uploading, but Meta did not do so).

seeder/downloader . . . from FB servers." Ex. 73, Meta_Kadrey_00120239; *see also* Ex. 74, Bashlykov 30(b)(1) Tr. Vol. II 286:15-287:22. By not using Meta IP addresses, "corporate computers," or Meta servers, Meta hoped its torrenting activity would be untraceable back to the company. Moreover, Meta's concerns about the risks of "seeding" data were clearly not obviated by the script it wrote, as it continued to voice these concerns long after the script was created. Ex. 34, Clark 30(b)(6) Torrenting Tr. 120:12-121:22.

Ultimately, Meta's own expert conducted an analysis of the data Meta acquired from these pirated sources. According to her report, Meta downloaded 267.4 TB of data across LibGen, Internet Archive, and Z-Lib from April through June 2024 alone. Ex. 72, Table 3. There is no dispute that this amount of data—comprising tens of millions of books and other literary works— also contains Plaintiffs' Books. Frederiksen-Cross identified 666 copies[26] of the Copyrighted Books in the datasets Meta downloaded, with each Book appearing at least once. Ex. 72, Table 2 & App'x C. When asked why she didn't replicate Meta's torrenting as part of her analysis in order to assess how much data Meta uploaded and/or seeded, Frederiksen-Cross replied: "It is my practice and policy not to do anything in my testing that could possibly either tarnish the reputation of the attorneys with whom I'm working or expose me to risk." Ex. 67, BFC Tr. 163:3-6.

IV.   META'S UNLAWFUL COPYING HARMED PLAINTIFFS

A. Stealing Books Harms Authors.

Meta's piracy campaign resulted in the download of pirated datasets that collectively contained multiple copies of every Copyrighted Book, totaling hundreds of copies.[27] Copyright infringement has long been harmful to the literary industry. Ex. 76, Spulber Opening Report, ¶ 191 ("From an economic perspective, downloading illegal, pirated copies of copyrighted works and then redistributing those same pirated copies en masse . . . risks reducing demand for those works in the legal market. Such conduct causes concrete harm to authors."); *see also Andy Warhol Found. for Visual Arts, Inc. v. Goldsmith*, 11 F.4th 26, 50 (2d Cir. 2021) ("That harm [the destruction of

---

[26] 576 copies for the movant Plaintiffs, and 666 copies across all Plaintiffs.

[27] Ex. 3; Ex. 72, App'x C.

the broader market if the copying Warhol engaged in were to become widespread] is . . . self-evident."), aff'd 598 U.S. 508 (2023). In 2017 alone, "US publishers lost approximately $315 million in sales because of online piracy."[28] Meta's rampant copying showed little regard for the rights of authors, an attitude echoed by Meta's Director of Business Development, AI Partnerships:

> Q. Do you not care whether Meta protects human creativity rather than exploits it in connection with its LLMs?
>
> . . .
>
> THE WITNESS: It's not something that I've given a lot of thought to; so my answer is I don't care at this point.

Ex. 91, Choudhury Tr. 309:18-310:2.

B. Meta Deprived Plaintiffs of the Opportunity To Sell and License Their Books.

Meta's actions denied Plaintiffs the opportunity to license their Copyrighted Books to Meta to train Meta's LLMs. All Plaintiffs expressed willingness to license their Copyrighted Books as LLM training data in exchange for appropriate compensation. Ex. 4. And while Meta initially pursued licensing, it stopped after it discovered that it could obtain copyrighted works, including Plaintiffs' books, for free from pirated sources. Ex. 52 ("[S]ome people on the team have mentioned they think we should spend more on licensing data for pre-training . . . . Melanie was pretty clear that she wasn't planning to license data right now. And it may be less relevant given that we are using libgen . . ?"); Ex. 20, Nho. Tr. 176:20–178:8 (describing the process by which Meta engineers compared works offered for licensing by publishers and whether they were available in existing datasets); Ex. 75, Meta_Kadrey_00093382 (spreadsheet listing titles from ▇▇▇ and ▇▇▇ and whether they were also in LibGen). As noted earlier, Meta employees cautioned against pursuing *any* licensing given the perceived impact on Meta's fair use defense. Ex. 52, at 135 ("[T]he problem is that people don't realize that if we license once [*sic*] single book,

---

[28] Davis, Cheryl L. & Kazi, Umair, *Piracy of Books in the Digital Age*, THE ROUTLEDGE COMPANION TO COPYRIGHT AND CREATIVITY IN THE 21ST CENTURY, 21.

---

we won't be able to lean into the fair use strategy. So we will have to drop all of the libgen and books datasets.").[29]

C. Meta Has Benefited Significantly from Acquiring and Training Llama on Pirated Books.

Once Meta started training its LLMs on massive amounts of pirated books and other copyrighted works, its engineers noticed that their performance significantly improved.[30] Accordingly, Meta engineers repeatedly flagged websites they knew contained pirated works. Ex. 82, Clark 30(b)(6) Tr. 167:17-19 ("they did classify LibGen and Sci-Hub as illegal, pirated websites in their opinion"); Ex. 55, Clark 30(b)(1) Tr. 164:1-10; Ex. 83, Edunov Tr. 77:19-78:6. Unsurprisingly, the engineers also believed pirated books were the most useful or "best" datasets in their possession. Ex. 25 ("LibGen is the most valuable dataset that we have so far."); Ex. 84, Meta_Kadrey_00233487 ("[H]igh quality long data . . . e.g. libgen scitech, fiction & non-fiction books are equally if not more useful for short context benchmarks as well as improving long context adaptation of the model."). Meta plainly attributed significant value to the copyrighted works it took for free: a windfall to Meta, but not for authors, who were paid nothing.

LEGAL STANDARD

Courts should grant summary judgment when "the movant shows that there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law." Fed. R. Civ. P. 56(a). A fact is "material" if it might affect the outcome of the case, and a dispute is "genuine" if a reasonable trier of fact could resolve the issue in the non-movant's favor. *Anderson*

---

[29] Meta's copying risks additional harms to Plaintiffs because the output of Meta's Llama models is directly based on the inputs. Ex. 76, ¶¶192-230; *see also* Ex. 21, LeCun Tr. 185:14–185:16 ("[Y]ou give [an LLM] a piece of text and then you train it to just reproduce the text on its output"); Ex. 77, Bell Tr. 93:16–93:19 ("[L]et's say I wanted to use an LLM to write a book for me; then I think in that case my intuition would be that obviously I want to train on a bunch of books."). The Llama models can produce output that directly copies significant portions of Plaintiffs' Books. Ex. 78, Lopes Suppl. Rebuttal Report (revised memorization experiment); Ex. 79, Lopes Rebuttal Report, Table 4 (Ungar experiments show continuations of Plaintiffs' books).

[30] Ex. 80, Meta_Kadrey_00232217 ("Libgen is essential to meet [state of the art] numbers across all categories . . . Without LibGen, research scientists within OneLLM believe it would not be able to achieve the OTA numbers the industry shows"); Ex. 81, Meta_Kadrey_00108008 (LibGen led to improvements on benchmarks); Ex. 24 ("Anna's Archive is full of long context books that they will desperately need for llama4 to have long context and there was a substantial bump in quality from libgen for similar reasons.").

---

v. Liberty Lobby, Inc., 477 U.S. 242, 248 (1986). Summary judgment is also appropriate against a party who "fails to make a showing sufficient to establish the existence of an element essential to that party's case, and on which that party will bear the burden of proof at trial." *Celotex Corp. v. Catrett*, 477 U.S. 317, 322 (1986).

Courts may also grant summary judgment against a non-moving party's affirmative defenses. Relevant here, courts grant summary judgment against infringing defendants' fair use claims where "undisputed facts . . . push [a] case squarely into the legal realm" and the "issues . . . are not ones of historical fact, intent, or factual prediction." *Thomson Reuters Enter. Ctr. GMBH v. Ross Intelligence, Inc.*, 2025 WL 458520, at *7 (D. Del. Feb. 11, 2025) (Bibas, J.); *see also BMG Music v. Gonzalez*, 430 F.3d 888, 890 (7th Cir. 2005) (affirming summary judgment against fair use defense); *Sony BMG Music Entm't v. Tenenbaum*, 672 F. Supp. 2d 217, 227 (D. Mass. 2009) (granting summary judgment against fair use defense).

ARGUMENT

V.   META COMMITTED DIRECT COPYRIGHT INFRINGEMENT AS A MATTER OF LAW

The Copyright Act protects a copyright owner's "exclusive rights to," *inter alia*, "reproduce the copyrighted work in copies[.]" 17 U.S.C. § 106. To establish direct copyright infringement, a plaintiff must demonstrate that the defendant violated one of the exclusive rights protected under the statute by showing "(1) ownership of a valid copyright, and (2) copying of constituent elements of the work that are original." *Range Rd. Music, Inc. v. East Coast Foods, Inc.*, 668 F.3d 1148, 1153 (9th Cir. 2012); *see also Skidmore as Tr. for Randy Craig Wolfe Tr. v. Led Zeppelin*, 952 F.3d 1051, 1064 (9th Cir. 2020) (same). The second prong "contains two separate components: 'copying' and 'unlawful appropriation.'" *Corbello v. Valli*, 974 F.3d 965, 973-74 (9th Cir. 2020). "Copying can be demonstrated either through direct evidence or by showing that the defendant had access to the plaintiff's work and that the two works share similarities probative of copying, while the hallmark of 'unlawful appropriation' is that the works share substantial similarities." *Id*. (cleaned up). When there is direct evidence of copying, "[a]

showing of 'substantial similarity'" between the copyrighted work and the allegedly infringing work "is irrelevant[.]" *Range Rd. Music*, 668 F.3d at 1154.

As there is no dispute that Plaintiffs own valid copyrights of the Copyrighted Books and undisputed direct evidence of Meta's copying, the Court should grant summary judgment to Plaintiffs on direct infringement due to Meta's copying the Books from pirated databases.

A. Plaintiffs Own Valid Copyrights for Each of Their Copyrighted Books.

"A requisite element of any claim for copyright infringement is ownership of the copyright at the time of the alleged infringement." *Rearden LLC v. The Walt Disney Co.*, 2024 WL 3956318, at *3 (N.D. Cal. Aug. 26, 2024).[31] Further, the Copyright Act provides: "[i]n any judicial proceedings the certificate of a registration made before or within five years after first publication of the work shall constitute prima facie evidence of the validity of the copyright and of the facts stated in the certificate." 17 U.S.C. § 410(c); *Aquarian Found., Inc. v. Lowndes*, 127 F.4th 814, 819 (9th Cir. 2025) (affirming reliance on registrations as prima facie evidence of plaintiff's authorship).

Here, each Plaintiff owns the registered copyright for their Copyrighted Books, obtained certificates of registration within five years after their first publication, and owned the copyrights at the time of Meta's infringement. Appendix A provides further information about the registrations owned by each Plaintiff, as do Exhibits 1 and 96, which show the registration certificates for the Copyrighted Books and select excerpts of the Copyrighted Books, respectively.

B. Meta Copied Plaintiffs' Copyrighted Books Without Permission.

Courts consistently hold that downloading copyrighted works without permission is direct infringement.[32] Here, the record establishes that Meta downloaded full, unauthorized copies of

---

[31] A plaintiff may be either the legal or beneficial owner, such that an author who transferred their exclusive rights under the Copyright Act in exchange for percentage royalties based on sales or license fees retains standing to sue to protect their economic interest in the copyright. *See DRK Photo v. McGraw-Hill Glob. Educ. Holdings, LLC*, 870 F.3d 978, 983, 988 (9th Cir. 2017).

[32] *See, e.g.*, *Napster*, 239 F.3d at 1014 ("Napster users who download files containing copyrighted music violate plaintiffs' reproduction rights."); *Columbia Pictures Indus., Inc. v. Fung*, 710 F.3d 1020, 1034 (9th Cir. 2013) ("downloading copyrighted material . . . violates the copyright holder's . . . right to reproduction"); *see generally* 4 Nimmer on Copyright § 13D.02 ("For instance, photocopying the entirety of a novel is plainly actionable.").

---

each Copyrighted Book, which directly infringed Plaintiffs' exclusive rights of reproduction. Specifically, Meta intentionally copied the pirated books datasets known as Books3, LibGen, Z-Lib, and Internet Archive ("IA"). Collectively, these pirated datasets contain full copies of all the Copyrighted Books, and Meta made at least *692* copies of the Copyrighted Books between the data it torrented and the training datasets it created. Ex. 3, Summary Table of Copies.[33]

Books3. It is undisputed that Meta downloaded the Books3 dataset and "used most of [it] as training data to train one or more Llama Models." Ex. 2, RFA No. 16; Ex. 85, Meta_Kadrey_00054894 (Meta employee explains "books3 is included in the pile, so I took it for free"). In April of 2023, a Meta employee created an inventory of the books it acquired via Books3, called "Books3 Titles." Ex. 39, Esiobu Tr. 34:18-35:11. That inventory shows that Meta's copy of Books3 supplied Meta with a copy of each of the Copyrighted Books. Ex. 86, Meta_Kadrey_0000250 (spreadsheet showing inventory of Books3, including each of Plaintiffs' Books). An inspection of the Books3 dataset[34] that Meta copied into its training data confirmed that Meta reproduced a full copy of the text of all but two[35] of the Copyrighted Books for use in training. Ex. 27, ¶ 162.

LibGen. Meta also admits to downloading a large cache of books from LibGen. Ex. 87, 30(b)(6) Bashlykov Tr. 42:11-43:14 (admitting Meta acquired multiple fiction and non-fiction books datasets from LibGen); *see also* Ex. 6, ROG No. 1 (identifying "https://libgen.is" as the website Meta used). Meta's LibGen dataset contains copies of over 758,000 pirated books, including full copies of the vast majority of the Copyrighted Books. Ex. 27, ¶¶ 147, 163. Meta employees routinely made copies of the LibGen dataset. *See e.g.,* Ex. 36 (copied as early as December 16, 2022); Ex. 88, Meta_Kadrey_00128310 (another copy downloaded by May 15,

---

[33] In this section, Plaintiffs refer to quantities of copies for solely the 11 movant Plaintiffs. The parties' expert reports quantify copies across all 13 Plaintiffs.

[34] For later use as training data, Meta combined the Books3 pirated dataset with another dataset of books called Gutenberg ("B3G"). The Gutenberg dataset is limited to public domain books.

[35] Meta produced the works it sourced from Books3, including a full copy of all of the Copyrighted Books except for *The Wrong Dead Guy*, which is missing 30 pages, and *The Golden Child*, which is missing entirely. Ex. 27, ¶ 162. There is, however, a full copy of *The Wrong Dead Guy* in the works that Meta sourced and produced from LibGen. *Id.* at ¶ 163.

2023); Ex. 17 (another copy downloaded by April 11, 2024).[36] Meta also reproduced the full text of the books obtained from LibGen in its training data, albeit only after removing the works' copyright management information ("CMI"). Ex. 27, ¶¶ 147, 163.

Z-Lib and Internet Archive. Meta further admits to downloading the Z-Lib and Internet Archive datasets via torrent. Ex. 72, ¶¶ 156-57 & Table 8. Meta downloaded 40 of the Copyrighted Books from each of Z-Lib and IA, collectively downloading 576 copies. Ex. 3. Each of the Books was copied at least one time from at least one of these datasets.

Evidence of just one copy of each Copyrighted Book would suffice to grant Plaintiffs summary judgment on direct infringement.[37] *Hachette v. Internet Archive* is instructive. There, book publishers brought a direct infringement claim against Internet Archive for "scanning print copies of the Works in Suit and lending the digital copies to users of [their] website without the [publisher's] permission." 664 F. Supp. 3d at 374 (S.D.N.Y. 2023). The court granted summary judgment to the book publishers on direct infringement where the defendant "violated the Publishers' reproduction rights[] by creating copies of the Works in Suit." *Id*. at 378 (citing 17 U.S.C. § 106(1)). Unlike the defendant in *Hachette*, Meta did not even buy the Copyrighted Books in the first instance. Ex. 2, RFA No. 9. Meta's large-scale copying of the Books without Plaintiffs' permission violated Plaintiffs' exclusive rights under § 106(1) and the Court should—as in *Hachette*—grant Plaintiffs partial summary judgment on direct infringement.

---

[36] Plaintiffs do not yet know how many copies Meta made as the Court denied Plaintiffs' efforts during this phase of discovery to identify all copies. *See* Dkt. 288 at 4-5; Dkt. 351 at 1-2.

[37] *See Hachette Book Grp., Inc. v. Internet Archive*, 664 F. Supp. 3d 370, 378 (S.D.N.Y. 2023), *aff'd*, 115 F.4th 163 (2d Cir. 2024) (defendant infringed plaintiffs' exclusive reproduction rights by copying entire books without permission); *see also Unicolors, Inc. v. Urb. Outfitters, Inc.*, 853 F.3d 980, 987 (9th Cir. 2017) (a district court may grant summary judgment even where the works are not identical, noting that "[a]llowing district courts to grant summary judgment for plaintiffs in copyright cases plays an important role in preserving the effect and weight of Rule 56"); *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 920, 926 (2d Cir. 1994) (no fair use where defendants photocopied entire articles, thereby engaging in "a systematic process of encouraging employee researchers to copy articles so as to multiply available copies while avoiding payment").

VI. META'S INITIAL REPRODUCTION OF PIRATED COPIES OF PLAINTIFFS' BOOKS IS NOT FAIR USE.

Fair use is typically adjudicated through a four-factor analysis prescribed by statute. 17 U.S.C. § 107. That analysis is a mixed question of law and fact. *E.g., SOFA Entm't, Inc. v. Dodger Prods., Inc.*, 709 F.3d 1273, 1277 (9th Cir. 2013). This mixed question often turns on the specific conduct of the infringer. 17 U.S.C. § 107; *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 23-25 (2021). For this reason, the fair-use analysis often requires jury findings and is inappropriate for full disposition at summary judgment. *See Los Angeles News Serv. v. KCAL-TV Channel 9*, 108 F.3d 1119, 1123 (9th Cir. 1997). Nevertheless, there exists a discrete set of infringing acts for which courts have held fair use cannot apply as a matter of law. Two such acts are relevant here. First, copying entire works from pirated databases to avoid compensating the rights holder cannot be fair use. That is unmitigated piracy not subject to a fair-use defense. Second, acquiring and sharing copyrighted works using P2P file sharing networks—a specific method of unmitigated piracy resulting in distributing copyrighted material to unknown third parties—also cannot be fair use. Meta engaged in both unprotected activities, warranting summary judgment to Plaintiffs.

A. Meta's Unmitigated Piracy

For hundreds of years, courts have held that unmitigated piracy of copyrighted works, i.e., the duplication of entire works to avoid compensating rightsholders, is not fair use. *See, e.g., Folsom v. Marsh*, 9 F. Cas. 342, 342-45 (C.C.D. Mass. 1841) (explaining "it is as clear, that if [defendant] thus cites the most important parts of the work, with a view, not to criticise, but to supersede the use of the original work, and substitute the review for it, such a use will be deemed in law a piracy") (Story, J.); *see, e.g., Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 550 (1985) ("As Justice Story's hypothetical illustrates, the fair use doctrine has always precluded a use that 'supersede[s] the use of the original.'"). That through-line has been applied to Internet piracy. *E.g., United States v. Slater*, 348 F.3d 666, 669 (7th Cir. 2003). The uncontroversial implication is that for fair use to apply, the work that was copied must have been lawfully acquired in the first place. *Atari Games Corp. v. Nintendo of Am., Inc.*, 975 F.2d 832, 843

(Fed. Cir. 1992) ("To invoke the fair use exception, *an individual must possess an authorized copy of a literary work.*") (emphasis added).[38]

Here, there is no dispute that Meta torrented over 267 TB of pirated copyrighted data, i.e., data from known illegal websites, nor is there any dispute that Meta acquired at least hundreds of thousands of additional pirated works via direct download from such sites. Those copies are all unmitigated piracy because they were made for the very purpose of acquiring the works without compensating the rightsholders. What's more, Meta knew that acquiring known pirated material could not be excused under the fair use doctrine. Recall Meta implemented a "mitigation" on its LibGen use where it would not use files whose titles self-identified as pirated or stolen. Ex. 90 at -740-41. If fair use protects unmitigated piracy of copyrighted works, even in the context of LLMs, then Meta would have had no reason to implement that mitigation in the first place.

Courts uniformly hold that allowing application of a fair use defense to this type of infringement is inimical to fair use because simply stealing a work in its entirety for financial benefit serves none of the goals of fair use. *See Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 578 n.10 (1994) (noting that "'most infringements are simple piracy,'" but "such cases are 'worlds apart from many of those raising reasonable contentions of fair use'") (quoting Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1132 (1990)); *see also Stewart v. Abend*, 495 U.S. 207, 236 (1990). When copyright infringement defendants attempted to appeal their conviction for the unauthorized download and dissemination of copyrighted software because the district court declined to give a fair use jury instruction, the Seventh Circuit held the defendants' fair use arguments "barely pass the straight-face test." *Slater*, 348 F.3d at 669. The court continued:

> Limited copying may be permissible for certain noncommercial, educational purposes, taking into account the nature of the copyrighted work and market considerations. These factors, however, weigh against application of the fair use doctrine to cases involving Internet piracy. [The defendants' Internet site] allowed members to obtain unlawful, digital duplicates of thousands of commercially

---

[38] *See also Perfect 10, Inc. v. Amazon.com*, 508 F.3d 1146, 1164 n.8 (9th Cir. 2007) (explaining Third Circuit and Federal Circuit "simply applied the general rule that a party claiming fair use must act in a manner generally compatible with principles of good faith and fair dealing" where defendants showed unauthorized movie clips and used source code accessed on false pretenses to replicate competitor's computer game, respectively) (citing *Harper & Row*, 471 U.S. at 562-63).

---

available software programs. The government also presented expert testimony on the harmful effect of Internet piracy on the potential market for the copyrighted work, though we think this point is fairly obvious. ***It is preposterous to think that Internet piracy is authorized by virtue of the fair use doctrine***.

Id. at 669 (emphasis added and citations omitted).

Other courts soundly reject the premise that infringement via online piracy can ever be legitimized as fair use. In *Tenenbaum*, 672 F. Supp. 2d at 234 (D. Mass. 2009), the court observed that digital downloading must be considered infringement, even if the owner of the copyrighted material fails to implement safeguards against digital piracy, because the "implications" of failing to do so "run exactly counter to the notion of fair use, which carves out an exception for uses that redound to the public's net benefit or do not reduce the incentives for creators." The court concluded that "[e]ncouraging piracy would do an immense disservice to the public purposes that animate copyright, with little commensurate gain." *Id.* This authority provides ample basis to hold as a matter of law that Meta's downloading pirated works for the express purpose of not paying for copyrighted material cannot be fair use; *see also infra* at B.3 (outlining how even if the Court did apply the fair use factors to such piracy, it would still fail as to P2P sharing as a matter of law).

B. Meta's P2P File Sharing

The additional element of P2P file sharing provides a separate, though related, basis to find as a matter of law that fair use does not apply to Meta's infringement. Courts, including the Ninth Circuit, hold that acquiring copyrighted works using P2P file sharing networks constitutes infringement without a valid defense. Indeed, Plaintiffs have not identified a ***single case*** in which a defendant successfully invoked fair use in relation to infringement by reproduction—whether downloading or uploading—of pirated works via P2P networks. *See e.g.*, *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 919 (2005); *Napster*, 239 F.3d at 1014-17 (users' downloading and uploading music files from the Napster platform is not fair use); *Gonzalez*, 430 F.3d at 890 (affirming summary judgment against fair use defense for P2P piracy); *Tenenbaum*, 672 F. Supp. 2d at 227 (granting summary judgment against fair use defense for P2P piracy). There

are no disputed facts that could make P2P file sharing of copyrighted material fair use. Because

Meta undisputedly copied Plaintiffs' Books in this manner, summary judgment is appropriate.

1. There Is No Genuine Dispute That Meta Acquired Millions of Pirated Copyrighted Works Using P2P File Sharing Networks.

Extensive and undisputed evidence shows that Meta torrented the entirety of several online

shadow libraries that each contain millions of pirated works. Meta's own expert determined that:

1) Meta torrented **267.4 TB** of data across Internet Archive (193.5 TB), LibGen's Non-Fiction database (10.3 TB), and Z-Lib (63.6 TB), Ex. 72, Table 3; and

2) Meta torrented **666 instances**[39] (i.e., copies) of Plaintiffs' works across those pirated datasets, *id.* at Table 2.

Further, Plaintiffs' experts concluded that:

1) Meta's source code contains hallmarks of torrenting LibGen and Anna's Archive via BitTorrent, including uploading pirated data, Ex. 89, Krein Report, ¶¶ 117-127, 158-168; and

2) Even with conservative estimates for the time that each torrent took to complete downloading, there is "a greater than 99.99999% chance that Meta uploaded at least one piece of Plaintiffs' works to another peer." Ex. 71, ¶¶ 20-30 & Table 1.

As discussed *supra*, torrenting involves the simultaneous download and upload of data in

a "tit-for-tat" bartering process. Though Meta admits it torrented a massive amount of copyrighted

data using up to six computers simultaneously over the course of more than *two months*,[40] its expert

maintains that "Meta took steps to prevent seeding data downloaded via BitTorrent, and these steps

should have prevented any distribution of Plaintiffs' works by Meta." Ex. 72, ¶ 100. Thus, she

concluded, "it is exceedingly unlikely that Meta seeded any of the Plaintiffs' works," *Id.* ¶ 134. At

its core, this opinion confirms that Meta *was* making copyrighted data available to other users after

it completed downloading each torrent. *See id.* ¶¶ 99-100. Indeed, Meta testified that even though

one of its data engineers wrote a program that would purportedly limit sharing files during this

"seeding" phase, managers within Meta still had concerns that seeding would occur and that it

could be traced back to Meta's servers. *See* Ex. 34, Clark 30(b)(6) Torrenting Tr. 120:12-121:7.

---

[39] 666 instances across all Plaintiffs, and 576 instances across the movant Plaintiffs. *See* Ex. 3.

[40] Ex. 34, Clark 30(b)(6) Torrenting Tr. 52:2-25.

Much more importantly, Meta's expert ignores uploading that Meta engaged in *during the downloading period itself*. BitTorrent's default configuration provides for continuous uploading during the "leeching" phase—simultaneous to downloading. As discussed *supra*, Meta used this default configuration. Ex. 67, BFC Tr. 41:1-22 & 44:8-45:15 (noting Meta engineers used default torrent configuration). In fact, the predominant risk of uploading arises during this phase, when the default operation of BitTorrent works around the one commonplace "network configuration" that Meta's expert identified as purportedly "block[ing] any connections not initiated by Meta." *See* Ex. 71, ¶¶ 8-12. Aside from that one network configuration that BitTorrent circumvents anyway, both sides' torrenting experts confirmed that Meta implemented *no other measures* to prevent uploading during the downloading or leeching period. Ex. 71, ¶ 19; Ex. 67, BFC Tr. 270:2-5 ("I don't know if anyone at Meta possessed that knowledge of libtorrent. I don't even know if they knew there were such parameters."). Under these facts, it is a virtual certainty that Plaintiffs' copyrighted material *was* shared with other users. *See* Choffnes Report, ¶¶ 20-30 & Table 1.[41]

2. Meta's Peer-to-Peer Sharing of Copyrighted Books Cannot Be Fair Use.

The illegality of downloading and sharing copyrighted material using P2P networks has been well-established for over two decades. *See, e.g., Grokster*, 545 U.S. at 919 ("[O]ne who distributes a device with the object of promoting its use to infringe copyright . . . is liable for the resulting acts of infringement by third parties.").[42] Without exception, courts have rejected attempts to employ a fair use defense to justify such infringement. *See id.*; *Glacier Films (USA),*

---

[41] In any event, whether another user actually downloaded the content that Meta made available is irrelevant. Meta "reproduced" the works as soon as it *made them available* to other peers. *Hotaling v. Church of Jesus Christ of Latter-Day Saints*, 118 F.3d 199, 203 (4th Cir. 1997) (an "offer to distribute the work" is "considered distribution"); *Warner Bros. Records v. Payne*, 2006 WL 2844415, at *3 (W.D. Tex. July 17, 2006) ("Listing unauthorized copies of sound recordings using an online file-sharing system constitutes an offer to distribute those works"). The Ninth Circuit credited *Hotaling* as creating a "'deemed distribution' rule" that would prohibit someone who possesses copyrighted material from "mak[ing it] available to the public," consistent with the prohibition against Napster users using the software "to make their collections available to all other Napster users." *Perfect 10,* 508 F.3d at 1162.

[42] *See also Sen. Hearing 108-920, Before Comm. on the Judiciary, U.S. Senate* (2003) (Register of Copyrights Marybeth Peters testifying that "the law is unambiguous. Using peer-to-peer networks to copy or distribute copyrighted works without permission is infringement, and copyright owners have every right to invoke the power of the courts to combat such activity. Every court that has addressed the issue agrees.").

Inc. v. Turchin, 896 F.3d 1033, 1043 (9th Cir. 2018) (dismissing as "baseless" the argument that "downloading or uploading of the copyrighted work" via a P2P network "was permitted by the doctrine of fair use"); *Napster*, 239 F.3d at 1014-17 (users' downloading and uploading files from Napster was not fair use).[43] For that reason, most courts, including this Court, treat the use of P2P networks to download or upload pirated IP as the open-and-shut case it is. *See In re DMCA*, 608 F. Supp. 3d at 879 ("In some cases, no analysis is required; it is obvious, for example, that downloading and distributing copyrighted music via peer-to-peer systems does not constitute fair use.") (citations omitted) (Chhabria, J.); *Gonzalez*, 430 F.3d at 890 (affirming summary judgment; download and retention of files from P2P network was not fair use, as "copiers such as [defendant] cannot ask courts (and juries) to second-guess the market and call wholesale copying 'fair use'") (Easterbrook, J.).

Indeed, courts have noted that—as a practical matter—*Grokster* precludes any ruling that P2P copying of pirated copyrighted material is fair use, regardless of any ultimate purpose or uses. *See e.g.*, *Tenenbaum*, 672 F. Supp. 2d at 227 ("*Grokster*'s secondary liability was premised on the fact that file sharing constituted a form of primary infringement, rather than a fair use."). Accordingly, the Court should follow this long line of authority and hold that Meta has no fair use defense for the Copyrighted Books that it undisputedly acquired using P2P file sharing networks.

3. Fair Use Cannot Protect Meta's "Use" of Uploading Copyrighted Material via P2P Sharing.

Even if the Court decides that the fair use analysis applies to Meta's unmitigated piracy and use of torrenting to obtain pirated copies of Plaintiffs' Copyrighted Books, it should nevertheless grant summary judgment to Plaintiffs under the four fair use factors regarding Meta's decision to make available to other P2P pirates millions of copyrighted books in exchange for faster download speeds. In determining whether a particular act of copying is a fair use, courts

---

[43] *See also Tenenbaum*, 672 F. Supp. 2d at 227 ("The defendant has offered the Court no legal authority that file sharing of the kind he engaged in constitutes fair use. In fact, a number of courts, including the Supreme Court, have found exactly the opposite.") (citations omitted); *In re Aimster Copyright Litig.*, 334 F.3d 643, 645, 655 (7th Cir. 2003) (users that "swap computer files containing [copyrighted] popular music" via internet-based services like Napster are "direct infringers") (Posner, J.).

consider (1) "the purpose and character of the use"; (2) "the nature of the copyrighted work"; (3) "the amount and substantiality of the portion used in relation to the copyrighted work as a whole"; and (4) "the effect of the use upon the potential market for or value of the copyrighted work." *Id.* The first and fourth factors are widely viewed as the most important, including for summary judgment. *Thomson Reuters*, 2025 WL 458520, at *7 (citing *Authors Guild v. Google*, 804 F.3d 202, 220 (2d Cir. 2015)). This is not a close call. There is nothing transformative about using pirated online databases to source free market substitutes of entire copyrighted works entitled to the highest copyright protection under the law, while simultaneously contributing "bandwidth, content, storage, and processing power" to other P2P pirates. Ex. 67, BFC Tr. 101:7-104:5. The use of piracy to further piracy can never be "fair use." *See Glacier*, 869 F.3d at 1043.

Factor 1: Meta's P2P Sharing Is Non-Transformative and Is Commercial.

Nothing about "the purpose and character of [Meta's] use" of pirated databases was "transformative," though it certainly was "commercial." *See Campbell*, 510 U.S. at 579. Meta torrented complete, unchanged copies of the Copyrighted Books and millions of other copyrighted works without paying for them. Meta did not reconfigure the default settings on its libtorrent client to prevent hosting or distributing data during the leeching or seeding phrases of torrenting. By failing to do that, Meta not only acquired millions of entire copyrighted works, including Plaintiffs', but it also ensured that others could acquire the works too. Reproducing or offering to distribute copyrighted works to anonymous peer-users is commercial because the peers receive works for free that they otherwise would have to buy.[44] The first factor thus clearly counsels against fair use. *See Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163, 190 (2d Cir. 2024) (rejecting fair use defense; "IA copies the Works in full and makes those copies available

_____

[44] *Napster*, 239 F.3d at 1015 ("[R]epeated and exploitative copying of copyrighted works, even if the copies are not offered for sale, may constitute a commercial use."); *id.* at 1018 (because "Napster users download a full, free and permanent copy of the recording," the "determination by the district court as to the commercial purpose and character of sampling is not clearly erroneous."); *Tenenbaum*, 672 F. Supp. 2d at 228 ("Peer-to-peer networks are based on an informal kind of exchange, whereby each user 'shares' his or her files and, at the same time, gains access to the files made available by others on the network") (citing 17 U.S.C. § 101 for the definition of "financial gain" as the "receipt, or expectation of receipt, of anything of value, including the receipt of other copyrighted works").

to the public in their entirety," which does not "achieve a transformative secondary purpose, but [] supplant[s] the originals"); *Napster*, 239 F.3d at 1015 ("Courts have been reluctant to find fair use when an original work is merely retransmitted in a different medium.").[45]

Factor 2: Plaintiffs' Copyrighted Books Are Fundamentally Creative.

The second factor, "the nature of the copyrighted work," 17 U.S.C. § 107(2), "calls for recognition that some works are closer to the core of intended copyright protection than others, with the consequence that fair use is more difficult to establish when the former works are copied." *Campbell,* 510 U.S. at 586. "Works that are creative in nature are 'closer to the core of intended copyright protection' than are more fact-based works." *Napster,* 239 F.3d at 1016 (quoting *Campbell*, 510 U.S. at 586). The Copyrighted Books are indisputably creative: the exact type of works that cut against a finding of fair use. *Cf. Marcus v. Rowley*, 695 F.2d 1171, 1176 (9th Cir. 1983) (noting that even factual works like instructional cooking booklets can contain creative aspects, and that the author's manner of assembly of facts "represented a creative expression").

Factor 3: Meta Took and Uploaded Complete Copies of the Copyrighted Books.

Under the third factor, courts consider whether "the amount and substantiality of the portion used in relation to the copyrighted work as a whole . . . are reasonable in relation to the copying." *Campbell,* 510 U.S. at 586-87. As discussed *supra*, Meta downloaded hundreds of copies of the Books *in their entirety* and made them available for upload as well. *See* Ex. 3; Ex. 27, ¶¶ 147, 163; Ex. 71, ¶ 30 ("In the case of Plaintiffs' works contained in Z-Library and [IA], [uploading by Meta] was nearly a foregone conclusion."). Meta thus did not "use" mere snippets of the Books. It copied them wholesale.

Factor 4: Meta's Conduct Causes Great Harm To Plaintiffs.

With respect to the fourth factor, "the effect on the market value for the copyrighted work," Meta's participation in a P2P network by making Plaintiffs' Copyrighted Books available for

---

[45] *See also UMG Recordings, Inc. v. MP3.Com, Inc.,* 92 F. Supp. 2d 349, 351 (S.D.N.Y. 2000) (transmittal of unauthorized copies of sound recording is an "insufficient basis for any legitimate claim of transformation"); *Infinity Broadcast Corp. v. Kirkwood,* 150 F.3d 104, 108 (2d Cir. 1998) (rejecting fair use for retransmitting copyrighted radio broadcasts over telephone lines); *Los Angeles News Serv. v. Reuters Television Int'l Ltd.,* 149 F.3d 987 (9th Cir. 1998) (rejecting fair use where TV news agencies copied copyrighted footage and retransmitted it to news organizations).

download by others (and potentially uploading hundreds of copies of those Books) interferes with (1) Plaintiffs' sales of their books across all mediums (physical and electronic), and (2) Plaintiffs' revenue opportunities in the well-established market for LLM training data. Proof of present or future harm is proof

> that the particular use is harmful, or that if it should become widespread, it would adversely affect the potential market for the copyrighted work . . . . *If the intended use is for commercial gain, that likelihood [of market harm] may be presumed. But if it is for a noncommercial purpose, the likelihood must be demonstrated.*"

Napster, 239 F.3d at 1016 (quoting *Sony Corp. of Am. v. Universal City Studios*, 464 U.S. 417, 451 (1984) (emphasis in original)).

Here, the Court may presume a likelihood of harm because Meta made the works available to users who benefited from access to pirated books without paying for them. *Napster*, 239 F.3d at 1018; *see also Tenenbaum*, 672 F. Supp. 2d at 228. Any "widespread" use of that sort naturally leads to market harm. Courts thus credit the obvious deleterious market effect of P2P sharing of copyrighted works. *E.g.*, *Hachette Book Grp.* 115 F.4th at 190 ("[IA] copies the Works in full and makes those copies available to the public in their entirety. . . . At least in this context, it is difficult to compete with free."); *Gonzalez*, 430 F.3d at 890 ("Music downloaded for free from the Internet is a close substitute for purchased music; many people are bound to keep the downloaded files without buying originals. . . . It is no surprise, therefore, that the only appellate decision on point has held that downloading copyrighted songs cannot be defended as fair use") (citing *Napster*, 239 F.3d 1004). Plaintiffs' economic expert, Daniel F. Spulber, notes that "downloading illegal, pirated copies of copyrighted works and then redistributing those same pirated copies en masse . . . to other users risks reducing demand for those works in the legal market," and "[s]uch conduct causes concrete harm to authors by diminishing demand for their copyrighted works." Ex. 76, ¶ 191.

CONCLUSION

For these reasons, Plaintiffs request that the Court grant Plaintiffs' motion for partial summary judgment.

Dated: March 10, 2025

By: /s/ *Maxwell V. Pritt*
     Maxwell V. Pritt

**LIEFF CABRASER HEIMANN &
BERNSTEIN, LLP**
Elizabeth J. Cabraser (SBN 083151)
Daniel M. Hutchinson (SBN 239458)
Reilly T. Stoler (SBN 310761)
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
(415) 956-1000
ecabraser@lchb.com
dhutchinson@lchb.com
rstoler@lchb.com

Rachel Geman (*pro hac vice*)
250 Hudson Street, 8th Floor
New York, New York 10013-1413
(212) 355-9500
rgeman@lchb.com

Kenneth S. Byrd (*pro hac vice*)
Betsy A. Sugar (*pro hac vice*)
222 2nd Avenue South, Suite 1640
Nashville, TN 37201
(615) 313-9000
kbyrd@lchb.com
bsugar@lchb.com

JOSEPH SAVERI LAW FIRM LLP
Joseph R. Saveri (SBN 130064)
Cadio Zirpoli (SBN 179108)
Christopher K.L. Young (SBN 318371)
Holden Benon (SBN 325847)
Aaron Cera (SBN 351163)

601 California Street, Suite 1505
San Francisco, California 94108
(415) 500-6800

jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
cyoung@saverilawfirm.com
hbenon@saverilawfirm.com
acera@saverilawfirm.com

Matthew Butterick (SBN 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
(323) 968-2632
mb@butterricklaw.com

BOIES SCHILLER FLEXNER LLP
David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsfllp.com

Maxwell V. Pritt (SBN 253155)
Joshua M. Stein (SBN 298856)
Margaux Poueymirou (SBN 356000)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsfllp.com
jstein@bsfllp.com
mpoueymirou@bsfllp.com

Jesse Panuccio (*pro hac vice*)
Jay Schuffenhauer (*pro hac vice*)
1401 New York Ave, NW
Washington, DC 20005
(202) 237-2727
jpanuccio@bsfllp.com
jschuffenhauer@bsfllp.com

Joshua I. Schiller (SBN 330653)
David L. Simons (*pro hac vice*)
55 Hudson Yards, 20th Floor
New York, NY 10001
(914) 749-8200
jischiller@bsfllp.com
dsimons@bsfllp.com

*Interim Lead Counsel for Individual and
Representative Plaintiffs and the Proposed Class*

**CAFFERTY CLOBES MERIWETHER &
SPRENGEL LLP**
Bryan L. Clobes (*pro hac vice*)
Alexander J. Sweatman (*pro hac vice*)
Mohammed A. Rathur (*pro hac vice*)
135 S. LaSalle Street, Suite 3210
Chicago, IL 60603
(312) 782-4880
bclobes@caffertyclobes.com
asweatman@caffertyclobes.com
mrathur@caffertyclobes.com

DICELLO LEVITT LLP
Amy Keller (*pro hac vice*)
Nada Djordjevic (*pro hac vice*)
James Ulwick (*pro hac vice*)
10 North Dearborn Street, Sixth Floor
Chicago, Illinois 60602
(312) 214-7900
akeller@dicellolevitt.com
ndjordjevic@dicellolevitt.com
julwick@dicellolevitt.com

David Straite (*pro hac vice*)
485 Lexington Avenue, Suite 1001
New York, New York 10017
(646) 933-1000
dsraite@dicellolevitt.com

**COWAN DEBAETS ABRAMS &
SHEPPARD LLP**
Scott J. Sholder (*pro hac vice*)
CeCe M. Cole
60 Broad Street, 30th Floor
New York, NY 10004
(212) 974-7474
ssholder@cdas.com
ccole@cdas.com

*Counsel for Individual and Representative
Plaintiffs and the Proposed Class*